



UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/607,370	06/30/2000	Reiner Kraft	ARC9-2000-0046-US1	8705

23334 7590 12/07/2004

FLEIT, KAIN, GIBBONS, GUTMAN, BONGINI
& BIANCO P.L.
ONE BOCA COMMERCE CENTER
551 NORTHWEST 77TH STREET, SUITE 111
BOCA RATON, FL 33487

EXAMINER

BURGE, LONDRA C

ART UNIT	PAPER NUMBER
----------	--------------

2178

DATE MAILED: 12/07/2004

Please find below and/or attached an Office communication concerning this application or proceeding.

Office Action Summary	Application No. 09/607,370	Applicant(s) KRAFT ET AL.	
	Examiner Londra C Burge	Art Unit 2178	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If the period for reply specified above is less than thirty (30) days, a reply within the statutory minimum of thirty (30) days will be considered timely.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

- 1) ☒ Responsive to communication(s) filed on 23 June 2004.
- 2a) ☒ This action is **FINAL**. 2b) ☐ This action is non-final.
- 3) ☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

- 4) ☒ Claim(s) 1-20 is/are pending in the application.
- 4a) Of the above claim(s) _____ is/are withdrawn from consideration.
- 5) ☐ Claim(s) _____ is/are allowed.
- 6) ☒ Claim(s) 1-20 is/are rejected.
- 7) ☐ Claim(s) _____ is/are objected to.
- 8) ☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

- 9) ☐ The specification is objected to by the Examiner.
- 10) ☐ The drawing(s) filed on _____ is/are: a) ☐ accepted or b) ☐ objected to by the Examiner.
Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).
Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).
- 11) ☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

- 12) ☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
- a) ☐ All b) ☐ Some * c) ☐ None of:
1. ☐ Certified copies of the priority documents have been received.
2. ☐ Certified copies of the priority documents have been received in Application No. _____.
3. ☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

- | | |
|------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------|
| 1) <input checked="" type="checkbox"/> Notice of References Cited (PTO-892) | 4) <input type="checkbox"/> Interview Summary (PTO-413)
Paper No(s)/Mail Date. _____ |
| 2) <input type="checkbox"/> Notice of Draftsperson's Patent Drawing Review (PTO-948) | 5) <input type="checkbox"/> Notice of Informal Patent Application (PTO-152) |
| 3) <input type="checkbox"/> Information Disclosure Statement(s) (PTO-1449 or PTO/SB/08)
Paper No(s)/Mail Date _____ | 6) <input type="checkbox"/> Other: _____ |

DETAILED ACTION

1. This action is responsive to communications: Amendment filed 6/23/2004.
2. Claims 1-20 are pending. Claims 1 and 14 are independent claims.
3. The rejections of claims 1-20 under 35 U.S.C. 103(a) as being unpatentable over Sanu and Adar have been withdrawn as necessitated by the amendment.

Claim Rejections - 35 USC § 103

4. **The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:**

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negated by the manner in which the invention was made.

5. **Claims 1-6, 10-20 are rejected under 35 U.S.C. 103(a) as being unpatentable over Raman U.S. Patent No. 6,249,794 B1 filed 12/23/1997 in view of King et al. (herein after King) U.S. Patent No. 6,161,114 filed 4/14/1999.**

In regard to independent claim 1, Raman discloses *retrieving a web document at an address, and extracting contents of the web document for rendering an intermediate dynamically constructed in-memory webpage representation of the web document at a hub processing unit which is formatted as if displayed for viewing in an end-use's web browser; loading secondary documents associated with the web document in order to render the secondary documents as part of the in-memory webpage representation, wherein the secondary documents include one or more images with textual content embedded therein* (Raman Col 1 Lines 58-64 and Col 7 Lines 38-51).

Raman does not specifically mention *analyzing and summarizing the in-memory webpage representation to produce a text map of the web page document of the textual contents therein; and using optical character recognition on the images to extract textual content for adding to the textual map for the webpage document*. However, King mentions a similar process (King Col 12 Lines 9-26). It would have been obvious to one of ordinary skill in the art to apply King to Raman, providing Raman the benefit of recognizing images and text beneficially to the analyzing and summarizing of the document.

In regard to dependent claim 2, Raman discloses *wherein the retrieving the web document at an address further comprises retrieving a document at an address selected from the group of addresses consisting of a nodal address, a network address, a URL and equivalents*. (Raman Col 5 Lines 22-31 and Col 16 Lines 55-67)

In regard to dependent claim 3, Raman discloses *wherein one or more images with textual content embedded therein include at least one of an in-line GIF images and an in-line JPEG image*. (Raman Col 5 Lines 11-22)

In regard to dependent claim 4, Raman does not specifically mention *wherein the loading secondary document further comprises the loading of secondary documents including one or more Java applets with textual content embedded therein*. However, King mentions a similar process (Kings Col 7 Lines 55-65). It would have been obvious to one of ordinary skill in the art to apply King to Raman, providing Raman the benefit of having a Java applet, which is compatible with most web documents.

In regard to dependent claim 5, Raman discloses *wherein the loading secondary documents further comprises the loading of secondary documents including web documents*

Art Unit: 2178

selected from the group of documents consisting of in-line frames, frames and equivalents
(Raman Col 43 Lines 35-40)

In regard to dependent claim 6, Raman does not specifically mention *wherein the loading secondary document further comprises the loading of secondary documents including one or more Java Script components with textual content embedded therein*. However, King mentions a similar process (Kings Col 7 Lines 55-65). It would have been obvious to one of ordinary skill in the art to apply King to Raman, providing Raman the benefit of having a Java script, which is compatible with most web documents.

In regard to dependent claim 10, Sanu teaches *wherein the tertiary sub-step of crawling further comprises the sub-steps of: issuing an HTTP command to a web server named in the URL; receiving contents of an HTML page as a result of the issued HTTP command; and passing on the contents of the HTML page to a Page Rendering subroutine*. (Raman Col 8 Lines 5-67 and Col 10 Lines 1-50)

In regard to dependent claim 11, Raman discloses receiving the contents of the HTML page in the Page Rendering subroutine; building an in-memory representation of a Layout for the HTML page and if more data is needed to properly form the representation, then performing the sub-steps of: requesting additional web-based information; gathering this additional web-based information; inserting any URLs associated with this additional web-based information into the second list and a URL cache (Raman Col 1 Lines 58-64 and Col 7 Lines 38-51 and Col 7 Lines 24-37); *building a final amended representation; and forwarding the final amended representation to an Extraction subroutine; wherein, if no more data is needed to properly form*

Art Unit: 2178

the in-memory representation, then forwarding the in-memory representation to the Extraction subroutine (Raman Col 1 Lines 58-64 and Col 7 Lines 38-51).

In regard to dependent claim 12, Raman discloses *accessing a set of memory structures of the Page Renderer; copying a text portion of the structures into a text map; inspecting any in-line GIF and JPEG image references in the memory structures; extracting alternate text attributes; adding the alternate text attributes to a text map (Raman Col 4 Lines 9-35); analyzing any In-line GIF and JPEG images; extracting text content from the GIF and JPEG images; adding text content from the images to the text map. (Raman Col 5 Lines 11-22)*

Raman does not specifically mention *invoking an optical character recognition engine; using the optical character recognition engine for text content and forwarding the text map to a Page Summarizer subroutine.* However, King mentions a similar process (King Col 12 Lines 9-26). It would have been obvious to one of ordinary skill in the art to apply King to Raman, providing Raman the benefit of recognizing images and text beneficially to the analyzing and summarizing of the document.

In regard to dependent claim 13, Raman discloses *receiving a text map from the Page Extractor subroutine; processing the text map in an application-specific manner; applying data extraction patterns to the text map; translating resultant data from the applying step; forwarding any URLS present in the text map to a manager subroutine; and forwarding any extracted data and metadata to application logic. (Raman Col 1 Lines 58-64 and Col 7 Lines 38-51)*

In regard to independent claims 14 and 20, claims 14 and 20 reflect similar subject matter claimed in claim 1 and is rejected along the same rationale.

In regard to dependent claim 15, claim 15 reflects similar subject matter claimed in claim 2 and is rejected along the same rationale.

In regard to dependent claim 16, claim 16 reflects similar subject matter claimed in claim 3 and is rejected along the same rationale.

In regard to dependent claim 17, claim 17 reflects similar subject matter claimed in claim 4 and is rejected along the same rationale.

In regard to dependent claim 18, claim 18 reflects similar subject matter claimed in claim 5 and is rejected along the same rationale.

In regard to dependent claim 19, claim 19 reflects similar subject matter claimed in claim 6 and is rejected along the same rationale.

6. **Claim 7-9 is rejected under 35 U.S.C. 103(a) as being unpatentable over Raman U.S. Patent No. 6,249,794 B1 filed 12/23/1997 in view of King et al. (herein after King) U.S. Patent No. 6,161,114 filed 4/14/1999 as claimed in claim 1, in further view of Meyerzon et al. (herein after Meyerzon 6,199,081) U.S. Patent No. 6,199,081 B1 filed 6/30/1998 issued 3/6/2001, and in further view of Meyerzon et al. (herein after Meyerzon 6,638,314) U.S. Patent No. 6,638,314 B1 filed 6/26/1998 issued 10/28/2003.**

In regard to dependent claim 7, Raman discloses *initializing a first list with seed values; checking if there are any URLs to be processed and, in response that any URL exists to be processed then performing the following steps of: determining if a URL is in a second list; and in response that a URL is not in the second list; then performing the following sub-steps of: inserting the URL into the first list (Raman Col 8 Lines 5-67); removing the URL from the first*

Art Unit: 2178

list after the scheduled crawling; entering the URL into the second list; and repeating the checking step until there are no more URLs to be processed; where if the determining step determines that the URL is in the second list then repeating the checking step until there are no more URLs to be processed. (Raman Col 8 Lines 5-67 and Col 10 Lines 1-50)

Raman does not specifically mention *scheduling the URL for crawling; crawling the URL when scheduled to do so*. However Meyerzon 6,199,081 B1 teaches *scheduling the URL for crawling* (Meyerzon 6,199,081 B1 Col 9 Lines 10-11) and *crawling the URL when scheduled to do so* (Meyerzon 6,199,081 B1 Col 9 Lines 20-21) and *scheduled crawling* (Meyerzon 6,199,081 B1 Col 9 Lines 10-11). It would have been obvious to one of ordinary skill in the art at the time the invention was made to apply Meyerzon 6,199,081 B1 to Raman, providing Raman the benefit of applying starting URLs that instruct the gather process where to begin as taught by Meyerzon 6,199,081 B1 Col 9 Lines 10-11.

Raman does not specifically mention *until there are no more URLs to be processed*. However Meyerzon 6,638,314 B1 teaches *of no more URLs to be processed* (Meyerzon 6,638,314 B1 Col 12 Line 57). It would have been obvious to one of ordinary skill in the art at the time the invention was made to apply Meyerzon 6,638,314 B1 to Raman, providing Raman the benefit of processing linked URLs within the filtered data corresponding to an electronic document to be complete as taught by Meyerzon 6,638,314 B1 Col 13 Line 58-59

In regard to dependent claim 8, Raman teaches *wherein the sub-step of initializing a first list with seed values further includes the list being a URL pool* (Raman Col 8 Lines 5-67 and Col 10 Lines 1-50)

Art Unit: 2178

Raman does not specifically teach of a URL pool. However, Meyerzon teaches of a *URL Pool* (Meyerzon Col 9 Line 34 i.e. list of all URLs). It would have been obvious to one of ordinary skill in the art at the time the invention was made to apply Meyerzon to Raman, providing Raman the benefit of applying a URL Pool, which maintains a list of URLs that are currently being processed or have not yet been processed, as taught by Meyerzon Column 9 Lines 22-24.

In regard to dependent claim 9, Raman teaches *wherein the sub-step of determining if a URL is in a second list further includes the second list being a visited pool*. (Raman Col 8 Lines 5-67 and Col 10 Lines 1-50)

Raman does not specifically teach of a visited pool. However, Meyerzon teaches a *visited pool* (Meyerzon Col 9 Lines 34-36 i.e. list of all URLs that have been visited). It would have been obvious to one of ordinary skill in the art at the time the invention was made to apply Meyerzon to Raman, providing Raman the benefit of applying a visited pool that contain a list of all URLs that have been visited or attempted to be visited during either the current Web crawl or a previous Web crawl.

Response to Arguments

7. Applicant's arguments with respect to claims 1-20 have been considered but are moot in view of the new ground(s) of rejection.

Conclusion

8. Applicant's amendment necessitated the new ground(s) of rejection presented in this Office action. Accordingly, **THIS ACTION IS MADE FINAL**. See MPEP § 706.07(a). Applicant is reminded of the extension of time policy as set forth in 37 CFR 1.136(a).

A shortened statutory period for reply to this final action is set to expire THREE MONTHS from the mailing date of this action. In the event a first reply is filed within TWO MONTHS of the mailing date of this final action and the advisory action is not mailed until after the end of the THREE-MONTH shortened statutory period, then the shortened statutory period will expire on the date the advisory action is mailed, and any extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of the advisory action. In no event, however, will the statutory period for reply expire later than SIX MONTHS from the date of this final action.

9. **The prior art made of record and not relied upon is considered pertinent to applicant's disclosure.**

Nehab et al.	U.S. Patent No. 6,029,182	issued	10/4/1996
Yamanaka et a.	U.S. Patent No. 5,983,247	issued	5/29/1997

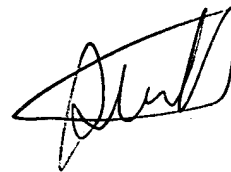
Any inquiry concerning this communication or earlier communications from the examiner should be directed to Londra C Burge whose telephone number is (571) 272-4122. The examiner can normally be reached on 8:30am to 5:00pm.

Art Unit: 2178

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Stephen Hong can be reached on (571) 272-4124. The fax phone number for the organization where this application or proceeding is assigned is 703-872-9306.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see <http://pair-direct.uspto.gov>. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free).

Londra C. Burge
11/17/04



STEPHEN S. HONG
PRIMARY EXAMINER